Australian Research Data Commons

Webinar: Library Carpentry

5 February 2019

PRESENTED BY

Chris Erdmann, Library Carpentry
Community & Development Director

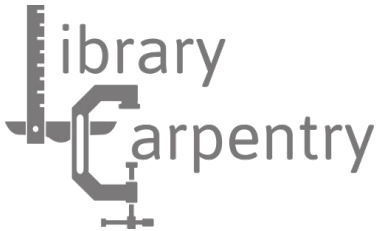
5 February 2019

NCRIS National Research Infrastructure for Australia
An Australian Government Initiative

The Carpentries: Teaching data science skills to researchers and people working in library- and information-related roles worldwide

Chris Erdmann, Library Carpentry Community and Development Director

February 5, 2019



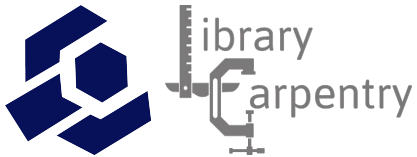
Hello, I'm Chris.

- CNET
- University of Washington
- Supreme Court of the US
- European Southern Observatory
- Harvard-Smithsonian Center for Astrophysics
- NC State
- **The Carpentries/CDL**
  - **Library Carpentry Community & Development Director**

library carpentry

Skills and perspectives to work with software and data are increasingly important as we generate more data.

With the emergence of our ability to generate increasing amounts of data, research and work in almost every domain has a data and computational component, including the whole new field of data science.



REALIZING THE POTENTIAL OF DATA SCIENCE

Final Report from the National Science Foundation Computer and Information Science and Engineering Advisory Committee Data Science Working Group

Francine Berman and Rob Rutenbar, co-Chairs
Henrik Christensen, Susan Davidson, Deborah Estrin, Michael Franklin, Brent Hailpern, Margaret Martonosi, Padma Raghavan, Victoria Stodden, Alex Szalay

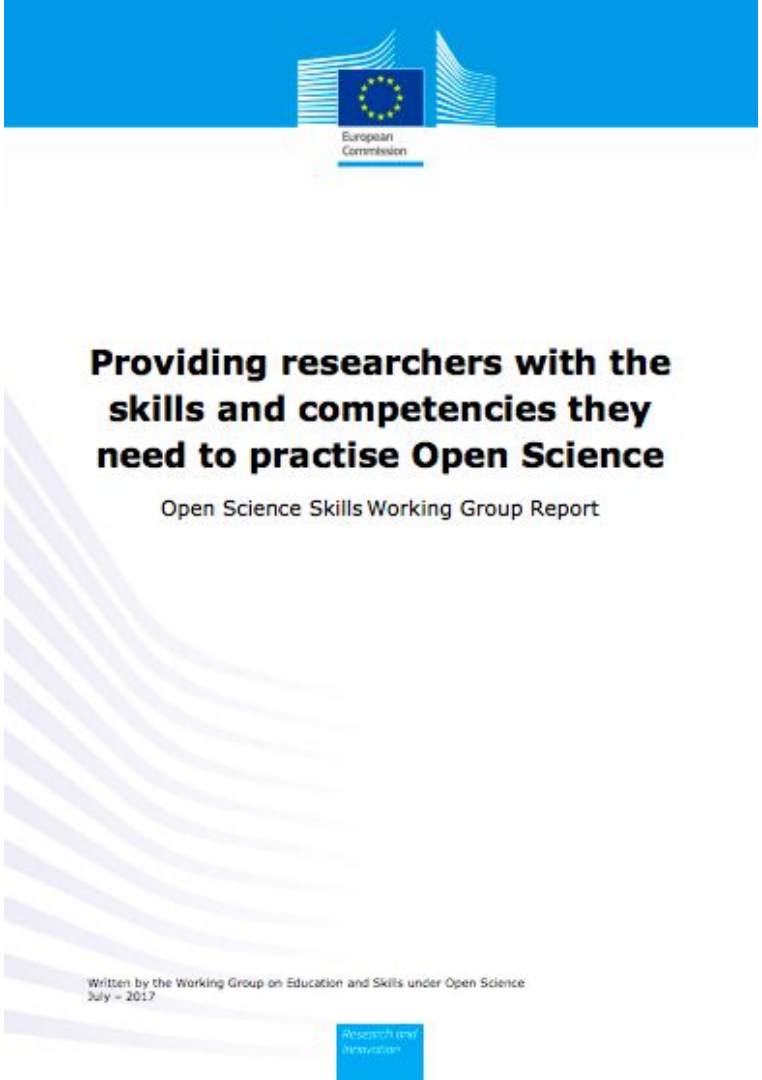
December 2016

library carpentry

The function of Federal advisory committees is advisory only. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the Advisory Committee, and do not necessarily reflect the views of the National Science Foundation.

library carpentry

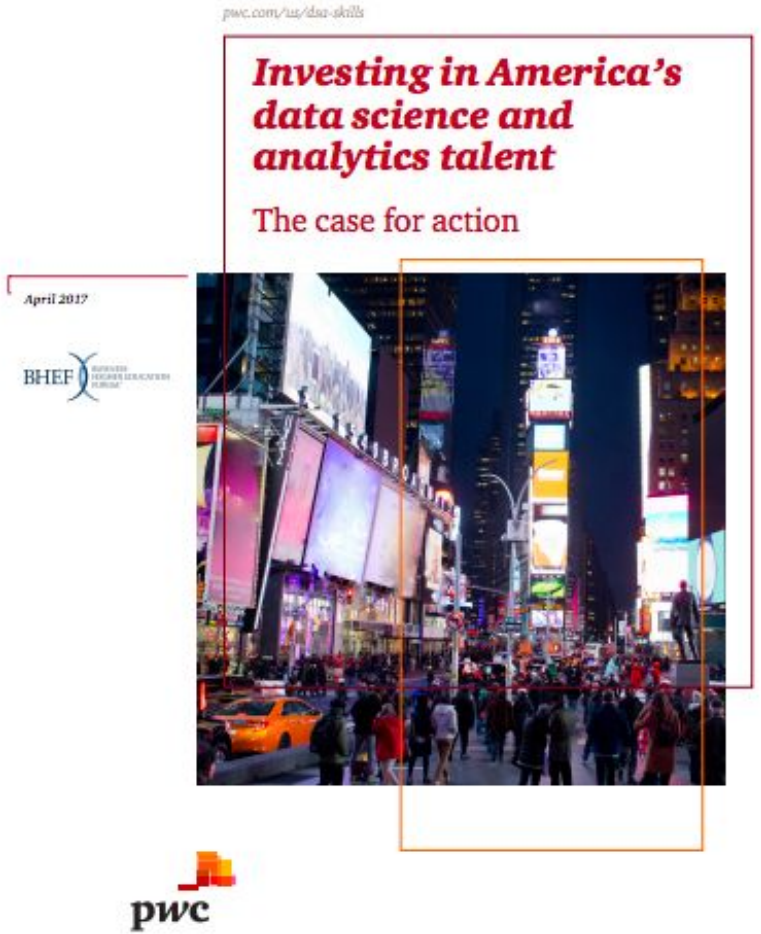
http://msdse.org/files/Creating_Institutional_Change.pdf



library carpentry

69% of business leaders in the United States will prefer job applicants with data skills by 2021.

23% of college and university leaders say their graduates will have those skills.



In August 2018, LinkedIn calculated that employers were seeking 151,717 more data scientists than exist in the U.S.

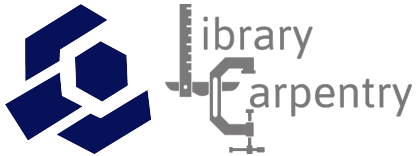
https://economicgraph.linkedin.com/resources/linkedin-workforce-report-august-2018

library carpentry

Rise of data science initiatives in academia

From the Data Science Community Newsletter by Noren & Stenger:

Brigham Young University, Caltech, Carnegie Mellon, College of Charleston, Columbia, Cornell, Dartmouth UMass, George Mason University, Georgetown University, Georgia Tech, Harvard, Illinois Wesleyan University, Johns Hopkins, Mid America Nazarene University, MIT, Northeastern University, Northern Kentucky University, Northwestern, Northwestern College in Iowa, Ohio State University, Penn State University, Princeton, Purdue, Stanford, Tufts University, UC Berkeley, UC Davis, UC Irvine, UC Merced, UC Riverside, UC San Diego, UCLA, UIUC, University of Iowa, University of Michigan, University of Oregon, University of Pennsylvania, University of Rochester, University of San Francisco, University of Warwick, University of Washington, UT Austin, UW Madison, Vanderbilt University, Virginia Tech, Washington University in St. Louis, Middle Tennessee State University, NYU, Amherst College, Brown, CU Boulder, Duke, Illinois Institute of Technology, Lehigh University, Loyola University - Maryland, Rice University, SUNY at Stony Brook, UC Santa Barbara, UC Santa Cruz, UCSF, UMass Amherst, UNC - Wilmington, University of Vermont, University of Arizona, University of British Columbia, University of Chicago, University of Virginia, USC, Worchester Polytechnic, Yale



70 and counting...

Importance of research software & training

- 92% of academics use research software
- 69% say that their research would not be practical without it
- 56% develop their own software (worryingly, 21% of those have no training in software development)

S.J. Hettrick et al, UK Research Software Survey 2014 [Data set]. Zenodo. http://doi.org/10.5281/zenodo.14809

Software
Sustainability
Institute

The Atlantic
@TheAtlantic

Follow

Forget Excel: This Was Reinhart and Rogoff's Biggest Mistake

Forget Excel: This Was Reinhart and Rogoff's Biggest Mistake

Correlation is not causation

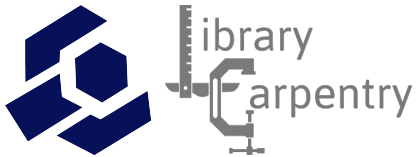
theatlantic.com

2:59 PM - 18 Apr 2013

Our path to better science in less time using open science tools

Reproducibility has long been a tenet of science but has been challenging to achieve—we learned this the hard way when our old approaches proved inadequate to efficiently reproduce our own work. Here we describe how several free software tools have fundamentally upgraded our approach to collaborative research, making our entire workflow more transparent and streamlined. By describing specific tools and how we incrementally began using them for the **Ocean Health Index** project, we hope to encourage others in the scientific community to do the same—so we can all produce better science in less time.

Lowndes, Julia S. Stewart, et al. "Our path to better science in less time using open data science tools." *Nature ecology & evolution* 1.6 (2017): 160.

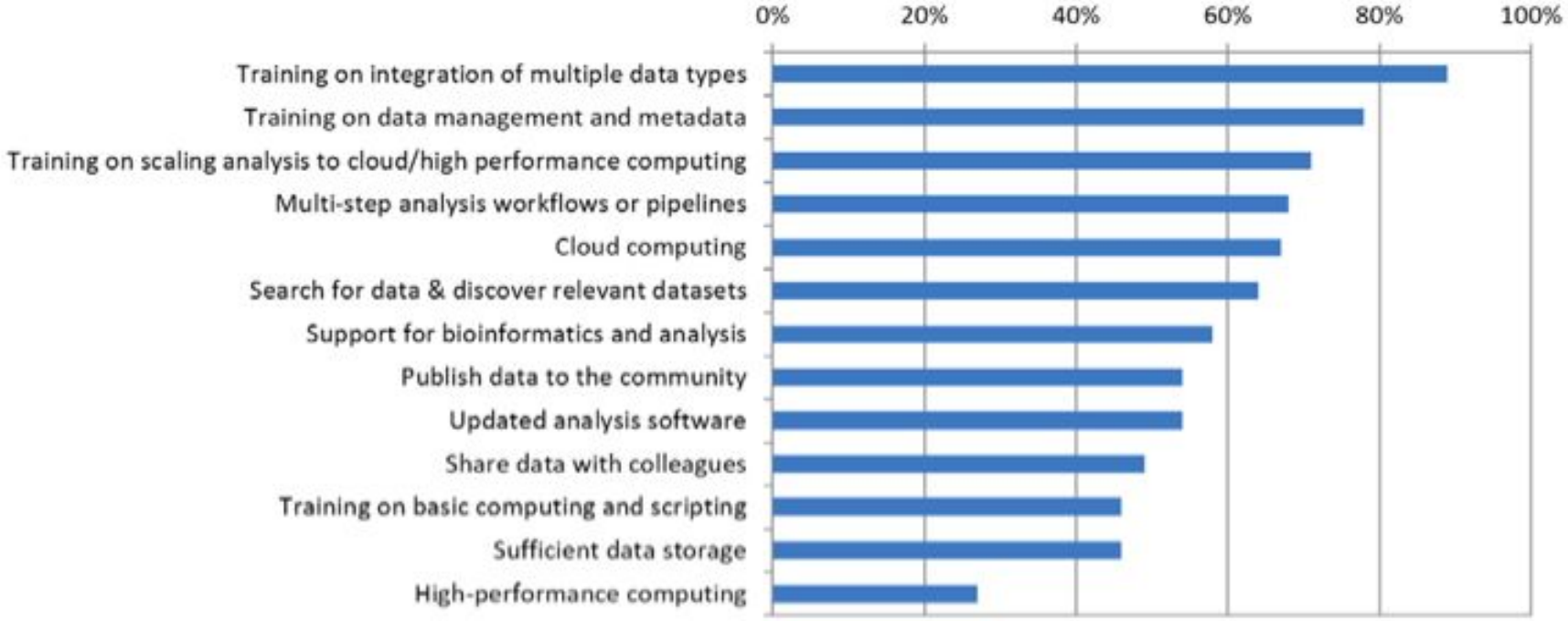


Researchers are very interested in learning these skills

Survey by Bioinformatics Resource Australia on what it would be most useful for them to offer

http://braembl.org.au/news/braembl-community-survey-report-2013

Current unmet needs



Barone L, Williams J and Micklos D. Unmet Needs for Analyzing Biological Big Data: A Survey of 704 NSF Principal Investigators (2017)

How do we scale data and software skills along with data production?

library carpentry

Building skills and community

- Creating training ‘in the gaps’ that is accessible, approachable, aligned and applicable
- Peer-led hands-on intensive workshops
- Volunteer instructors
- Open and collaborative lesson materials
- Creating and supporting community

THE CARPENTRIES

Non-profit organization that:

- Trains people in software development and data science skills for more effective work and career development
- Builds community and local capacity for teaching and learning these skills and perspectives

Note: “Carpentry” means “the basics” like learning how to nail two boards together or put up a wall straight.

library carpentry

Workshops

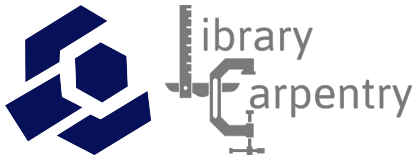
- 2-days, active learning
- Feedback to learners throughout the workshop
- Trained instructors
- Friendly learning environment

Focus

Data Carpentry
Domain-specific, research data-related

Software Carpentry
Domain agnostic, research workflow/software-related

Library Carpentry
Library and information/workflow-related, Carpentries onboarding, community outreach and advocacy-driven

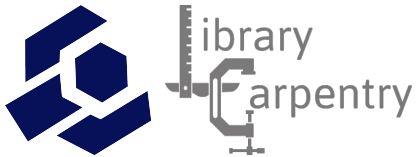


Lessons

Software Carpentry
Command line, version control, programming

Data Carpentry
Ecology, Genomics, Geospatial, Social Science, Atmospheric Science

Library Carpentry
Data introduction, command line, version control, data wrangling



Workshop goals

- Teach skills
- Get people started and introduce them to what's possible
- Build confidence in using these skills
- Encourage people to continue learning
- Positive learning experience

library carpentry

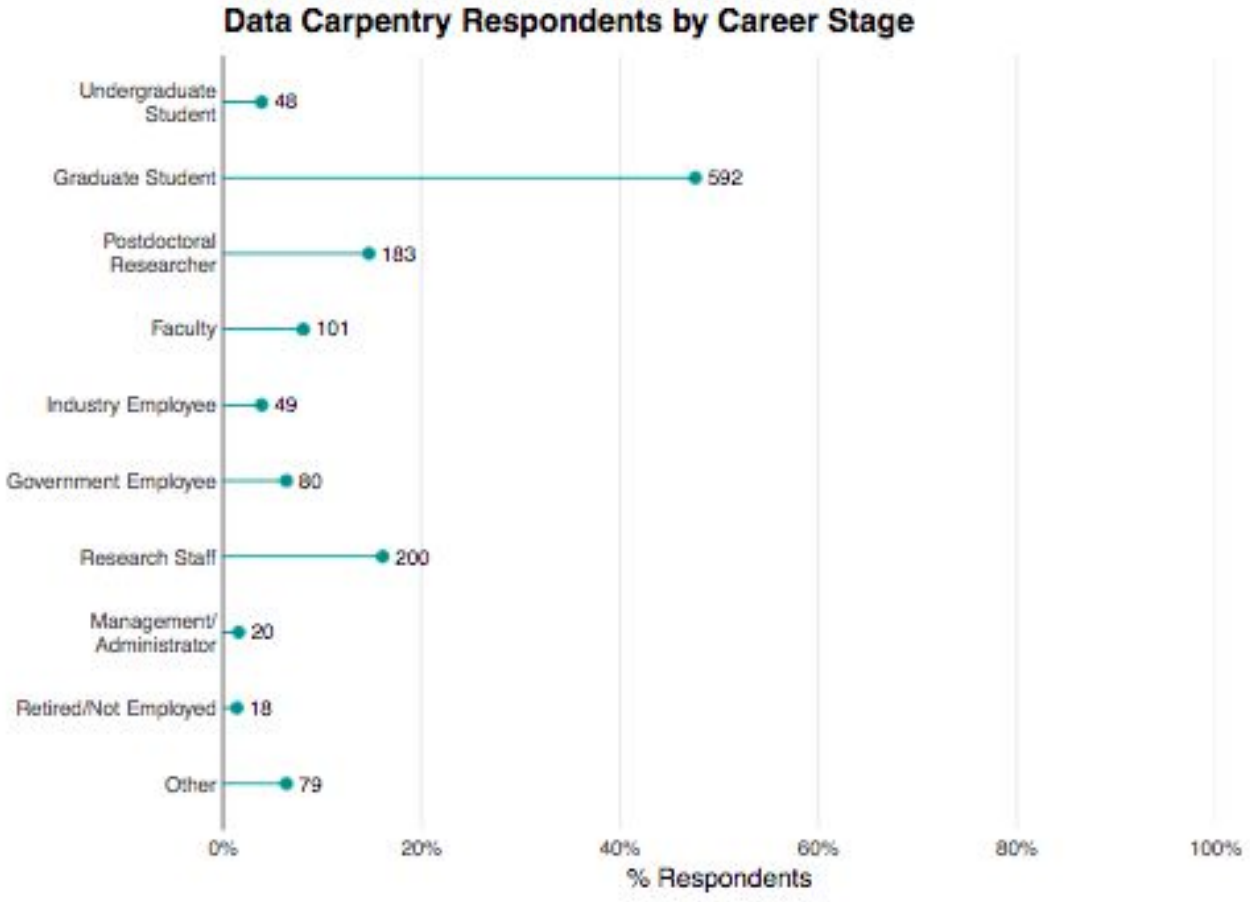
Our Workshops.
Our learners.

The Carpentries 2018 Annual Report
https://carpentries.org/files/assessment/TheCarpentries2018AnnualReport.pdf

Workshops worldwide

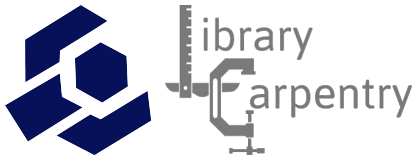
n = 1,350 on all 7 continents yes, even Antarctica

Who takes workshops?



66% of the Data Carpentry workshop attendees are early career.

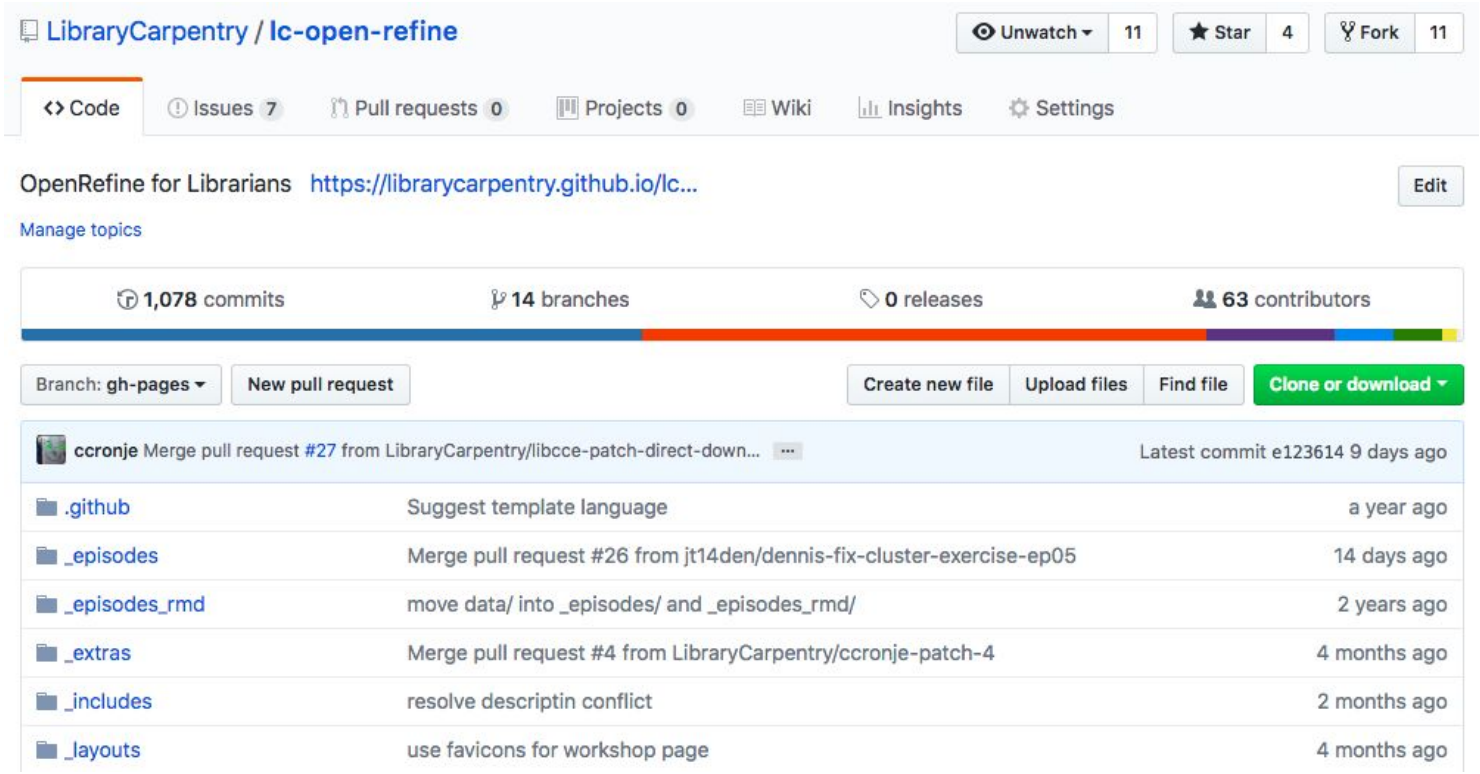
Analysis of Software and Data Carpentry's Pre- and Post-Workshop Surveys
https://doi.org/10.5281/zenodo.1325463



Instructors

Educational pedagogy the focus of Instructor training program: 2-days plus 3 other steps (edit a lesson, 1-hr discussion, demo). http://carpentries.github.io/instructor-training/

Open, collaboratively developed lessons



Community

A group of people excited about software and data skills and about sharing them with others

- Mentoring program and instructor onboarding
- Discussion groups and community calls
- Email lists
- Teaching at other institutions

library carpentry

Outcomes

Short and long term surveys show that people are learning the skills, putting them into practice in their work and have more confidence in their ability to do computational work.

The tools I learned in my Carpentry workshop:

"helped me to reshape my workflow into a far more efficient and robust process."

"are improving my ability to share data and code."

"helped facilitate my understanding of the problems and solutions to accessing and transforming data."

"[are] useful tools for training my own team."

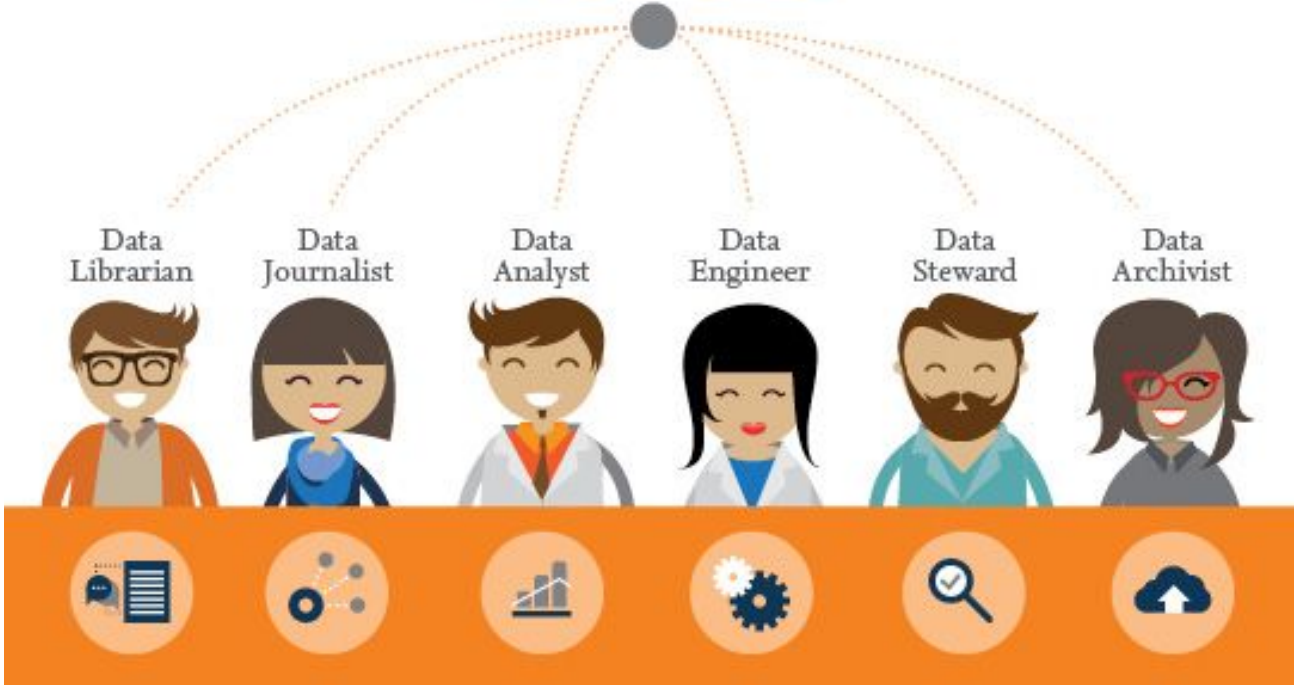
Figure: Perception of Workshop Impact

https://carpentries.github.io/assessment/learner-assessment/archives/2018/code/2018_January_long_term_report.html

Many lessons, and organizations, flowing together, forming The Carpentries

23 Research Data Things allowed thousands of librarians to familiarize themselves with research data topics. Library Carpentry aims to do the same for software and data skills.

DATA SCIENCE ROLES



https://libraryconnect.elsevier.com/articles/learning-about-research-data-lab-pitt-ischool

Growing community

The New England Software Carpentry Library Consortium (NESCLiC)

UMASS AMHERST LIBRARIES
www.library.umass.edu

MOUNT HOLYOKE
MOUNT HOLYOKE COLLEGE

Tufts
UNIVERSITY

Yale

VE RI TAS

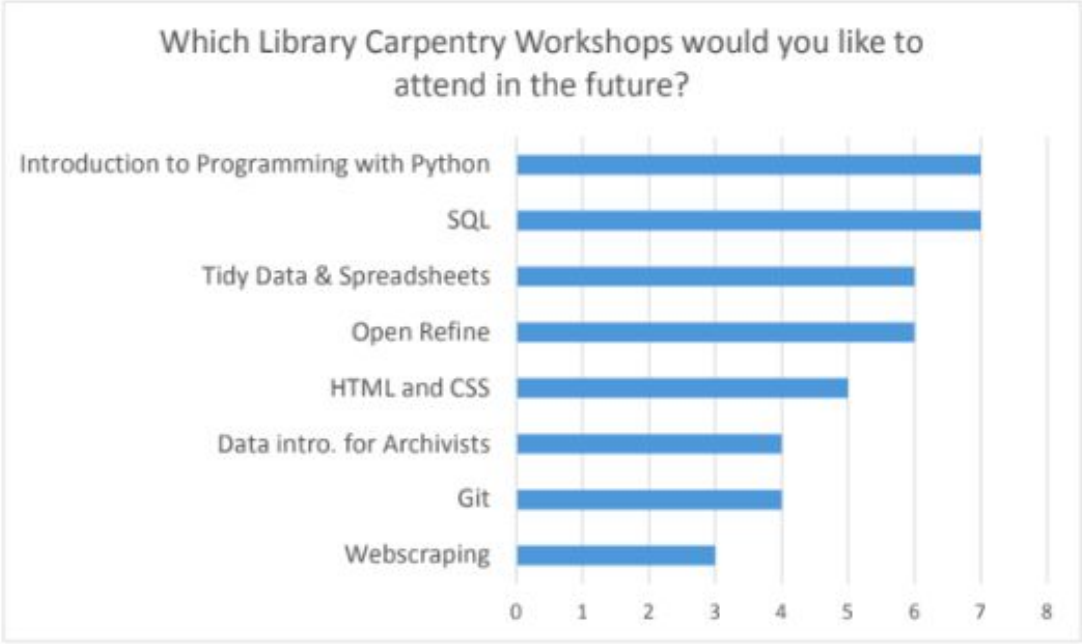
HARVARD
LIBRARY

DARTMOUTH

BROWN

library carpentry

NLM Library Carpentry Pilot Report



The Strategic Value of Library Carpentry and The Carpentries to Research Libraries

By Elaine L. Westbrooks

Instructors

Katrin Leinweber, Angelina Kraft, Konrad Förstner, Martin Hammitzsch, Luke Johnston, Mateusz Kuzak

Helpers

Chris Erdmann

General Information

This workshop aimed to train junior scientists in implementing the FAIR principles for research data & software management & development. We want to help you identify similarities and differences between these two scientific objects and apply respectively appropriate good practices in preparing, publishing and archiving your work.

It was a new, experimental workshop format that contextualises the highly practical lesson material from the Software and Data Carpentries with the FAIR principles

library Carpentry

Australian Research Data Commons

Top 10 FAIR Data & Software Things

about github repository download/cite #top10fair

Oceanography

Research Software

Research Libraries

Research Data Management Support

International Relations

library carpentry

https://librarycarpentry.org/Top-10-FAIR/

How can I get started? Contribute to a lesson.

Data intro for librarians
An introduction to data structures, regular expressions, and computing terms

Unix Shell
An introduction to command line interfaces and task automation using the Unix shell

OpenRefine
An introduction to cleaning up and enhancing a dataset using OpenRefine

Git Intro for Librarians
An introduction to version control using Git and GitHub for collaboration

SQL for Librarians
An introduction to relational database management using the SQLite tool

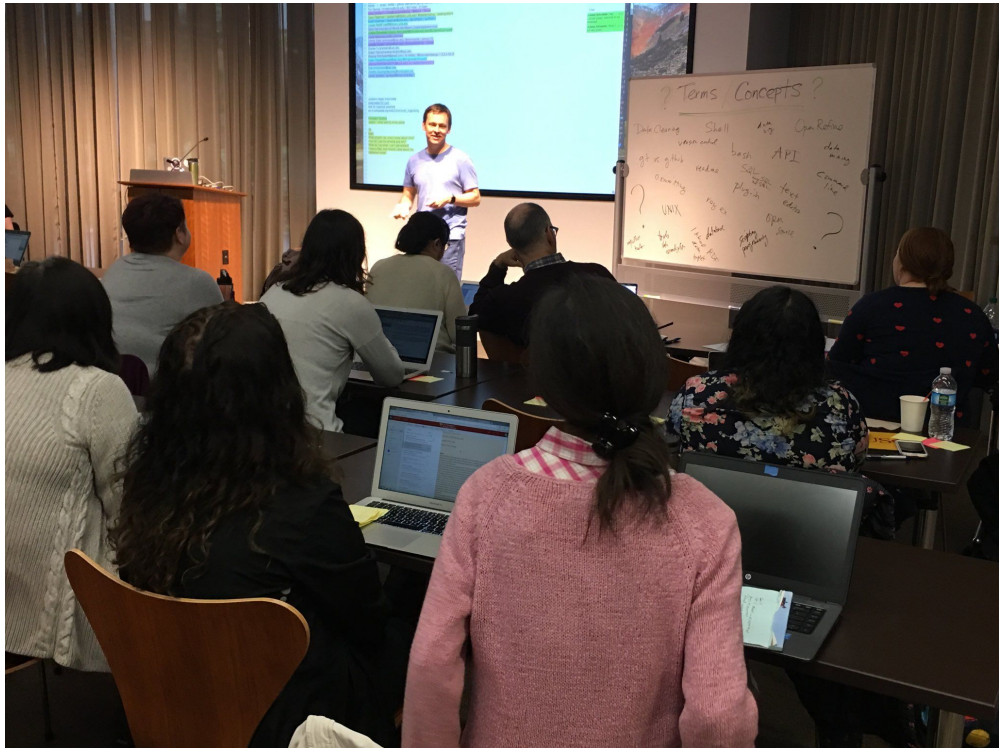
Webscraping
An introduction to extracting structured data from websites suing a range of tools

Tidy data for librarians
An introduction to good data organisation, which is the foundation of much of our day-to-day work in libraries.

Introduction to Python
An introduction to Python, a general purpose programming language

Data Intro for Archivists
An introduction to data structures, regular expressions, and computing terms for archivists

How can I get started? Host, Help, Teach.



How can I get started? Become a member.

	Bronze	Silver	Gold	Platinum
# of Coordinated Workshops	2	4	6	negotiable
Discount for additional coordinated workshops	20%	33%	50%	negotiable
Self-organized workshops at member organisation **	no-charge	no-charge	no-charge	no-charge
Number of instructors trained ***	0	6 online	15 with possibility for in-person^ training event	negotiable
Seat on the Carpentries Member Council	No	Yes	Yes	Yes
Train an in-house instructor trainer at member org	No	No	No	Available
Lesson development services	No	No	No	Available
Membership Dues (annual)	$5,000	$7,500	$15,000	Contact us

38 mentions of the Carpentries as an example and recommendation

Shifting to Data Savvy: The Future of Data Science In Libraries

Matt Burton
Liz Lyon
Chris Erdmann
Bonnie Tijerina

Thank you. Questions?

chris@carpentries.org

@libcce